

THE INSTANTIATION OF OMNIPAPER RDF PROTOTYPE IN THE CONTEXT OF SCIENTIFIC PUBLICATIONS

Teresa Susana Mendes Pereira

Polytechnic Institute of Viana do Castelo

Superior School of Business Studies

Valença, Portugal

tpereira@esce.ipvz.pt

Ana Alice Baptista

University of Minho

School of Engineering

Information System Department

Guimarães, Portugal

analice@dsi.uminho.pt

Abstract

Purpose of this paper

The purpose of this paper is to present an instance of the system developed in the OmniPaper project, regarding the mechanisms of distributed information retrieval. These mechanisms were developed for newspapers' articles and they were then instantiated in the context of the scientific publication. Another goal concerns the use of a central metadatabase developed to accomplish the syndication of contents, through the RSS approach.

Design/methodology/approach

One of the steps of the system's development was the definition of the metadata layer that supports the research and the navigation functionalities as well as the contents' syndication. Several tasks were performed for the definition of the metadata layer, namely: (1) analysis of several metadata standard vocabularies; (2) Selection of the metadata elements; (3) Definition of an application profile and the RSS template; (4) Development of a metadatabase, through the use of a native RDF database management system to store the RSS descriptions of the scientific publications; (5) Implementation of the search and navigation processes developed in the prototype through the use of the RDFS version of the WordNet and the RDFS version of classification system of Association for Computing Machinery Computing Classification System (ACM CCS); finally (5) Tests and validation of all developed functionalities.

Findings and value

The OmniPaper system can be instantiated to other domains other than news published in newspapers. The RSS technology is well suited for handling the description of scientific contents. RDF records that were used in the

OmniPaper RDF prototype were replaced by RSS. The subject and lexical thesauri were kept. This strong metadata layer allows the creation of several services that facilitate the conceptual search of scientific contents.

Originality and value of paper

This paper presents a system that uses a central metadatabase to support conceptual searching mechanisms. The metadatabase consists of RDF triples generated from: (1) RSS files that were, by their turn generated from OAI-PMH harvested metadata records; (2) a controlled vocabulary (ACM-CCS) implemented in RDF Schema and (3) an RDF version of WordNet. This is a solution for a value-added service for the scientific community that is fully based in state-of-the-art standard technologies and is fully open for integration with other systems. Moreover this could be implemented by journals to improve the current mechanisms used to access, distribute and disseminate the scientific research developments.

Research limitations/implications (if applicable)

The system implemented was tested but not evaluated in a real environment with specific users.

Keywords: Scientific publications, scholarly communication, metadata, information retrieval, RDF, WordNet and RSS.

Type of Paper: Research paper

1 Introduction

Today the Web is an important and widely used information source. In fact the increased use of the Web associated to the constant evolution of technologies has promoted the development of sophisticated information systems to facilitate the access and dissemination of scientific contents produced by scientific communities. However, the Web provides several information sources and, consequently, the identification of new contents or updates demands time. In fact, users spend a lot of time tracking a set of information sources to check for new contents or updates and sometimes some of the resources are not even accessed.

RSS is an XML-based format to syndicate information, or metadata. The content syndication helps the user to be up to date of new contents published in different and distributed information sources and improves the visibility of the contents published, guaranteeing that the user becomes aware when new contents are published.

In the context of scientific contents the use of the RSS technology introduces important advantages in the distribution and dissemination process of the results produced by scientific communities. In spite of the RSS being primarily used in relaying the latest entries' headlines of the newspapers and weblogs, it has been adapted to a wide range of uses in the description of Web contents, to enhance the rapid dissemination of the contents. Some journals already use the RSS format in the description of the scientific articles published, such as D-Lib, Ariadne, Nature Publishing Group's (NPG), et cet. In fact the syndication of metadata information of scientific contents has been contributing to the transformation of the current communication processes and information retrieval systems.

In the OmniPaper project sophisticated mechanisms of distributed information retrieval were investigated and developed in order to facilitate the access and

distribution of the news contents published in the newspapers. These functionalities were supported by a semantic metadata layer. To take advantage of the search and browsing functionalities developed in the OmniPaper project, an instance of the OmniPaper RDF prototype was created in the context of scientific publication. In fact the whole RDF structure (described below) developed in the OmniPaper was used in the context of scientific contents in order to provide a mechanism to facilitate the distribution and dissemination of scientific research developments. The metadata layer followed an RSS approach in order to enable the syndication of the metadata information of the scientific contents.

This paper intends to describe the research work conducted in the implementation of the instantiated system of the OmniPaper RDF prototype and deepens each step shown below.

2 The RDF Prototype Developed in the OmniPaper Project

The OmniPaper (Smart Access to European Newspapers, IST-2001-32174) was a project from the European Commission IST (Information Society Technologies) program. The OmniPaper project aims to (1) find and test mechanisms for retrieving information from distributed sources in an efficient way; (2) Find and test ways for creating a uniform access point to several distributed information sources; (3) Make this access point as usable and user-friendly as possible; (4) Lift widely distributed digital collections to a higher level. The purpose of the system implemented in the OmniPaper project was to enable users (professional journalists and occasional users) to have a simultaneous and structured access to the newspapers articles of a large number of digital European news providers. The OmniPaper was not a project about digitalization of news, but about bringing digitized news originating from various sources (and various formats) together (yaginuma et al, 2004).

One of its main aspects is the whole metadata layer of the system describing the metadata information of each local archive of digital news providers. This approach enables the user to search and navigate on the metadata information instead of performing integral text search, as it is usual in the most general information retrieval system. The OmniPaper architecture has two metadata layers: the Local Knowledge Layer and the Overall Knowledge Layer. The Local Knowledge Layer is composed of distributed metadatabases containing standard semantic descriptions of all the existent articles provided by each digital news provider; this enables a structured and uniform access to the available distributed archives. The Overall Knowledge Layer includes a conceptual layer that facilitates cross-archive browsing and navigation through the use of a web of concepts. Furthermore, this layer is enriched with multilingual and personalization functionalities for the interface with the user (Paepen, 2005).

The steps followed in the development of the OmniPaper RDF prototype (Baptista, 2004) were:

(1) Definition and development of the metadatabase. The definition of the metadata element set (the OmniPaper Application Profile) (Pereira et al, 2003b) to be used in the descriptions of the articles was based on the Dublin Core Metadata Terms (DCMT) [2]. This choice stems from its rich semantic interoperability; it is an ISO standard (15836:2003) [5]; it is an ANSI/NISO standard (Z 39.85-2001) [6]; it has been a stable element set since 1996; it has been a Dublin Core Metadata Initiative (DCMI) [3] recommendation since 1999 with its version 1.1; and finally due to the fact that it is a widely used metadata element set across boundaries of disciplines or application domains

(Baptista, 2004). However the DCMT didn't contain all the necessary elements in the description of the news articles. Therefore a new namespace RDF Schema (called "omni") had to be defined in order to add the metadata elements established by the consortium partners (Pereira et al, 2003b).

(2) Definition and development of the conceptual layer (subject + lexical thesaurus). The goal of the conceptual layer was to have an ontology-based web of concepts linked to the digital newspaper articles (Baptista, 2004). The used solution was the International Press Telecommunications Council Subject Codes (IPTC SC) [8]. The IPTC created and maintains a set of topics to be assigned as metadata values to news resources. The IPTC SC are currently split into 30 individual sets of topics usually relate to a specific area and are likely to be used exclusively in a specific metadata element of a news exchange format [9]. The IPTC SC is composed of a hierarchical three-level tree of subject codes, which describes the content of a set of concepts and has no associative relationship between its concepts. Topics of level Subject provide a description of the editorial content of news, a SubjectMatter provides a description at a more precise level and finally a SubjectDetail at a more specific level. In order to illustrate these hierarchy is presented an example in the followed table:

Subject		SubjectMatter		SubjectDetail	
04000000	Economy, Business & Finance				
		04003000	Computing & Information Technology		
				04003001	Hardware
				04003002	Networking
				04003003	Satellite technology
				04003004	Semiconductors & active components
				04003005	Software
				04003006	Telecommunications Equipment
				04003007	Telecommunications Services

Table 1. Example of hierarchical structure of the subject 'Economy, Business & Finance', one subject matter and a set of subjects detail included in this specific subject. Adapted from [8].

An ontology language provides semantics for a set of concepts and relations in order to produce the qualified and the possible interpretations. Some languages allow the definition of axioms or logical relations between terms for the same purpose [Mika, 2002].

The selection of ontology language is based on the type of knowledge structures intended to represent. Therefore, in the scope of the OmniPaper project, particularly on the RDF approach, it was developed a deep studied of several ontology language that could best fit the description needs, especially to achieve the description purpose of the hierarchical tree represented in the IPTC Subject Codes. The ontology languages analysed were based on languages associated with the Semantic Web, such as the Web Ontology Language (OWL), DARPA Agent Markup Language (DAML), Ontology Inference Layer", or "Ontology Interchange Language (OIL) and the Resource Description Framework Schema (RDF-S).

The analysis of the main features of the expressive Semantic Web languages contributed to the selection of the RDF Schema language to complete the representation of the hierarchical tree represented in the IPTC Subject Codes. This choice was based on the fact that the IPTC SC structure is not semantically that rich and due to the simple hierarchy of the concepts and relations presented in the IPTC SC structure, it was only necessary to define its hierarchical concepts. The use of a more expressive language wouldn't accomplish any beneficial use (Pereira et al, 2004).

The IPTC SC was then included in a metadatabase. The connection with the subject elements included in the hierarchical tree of the IPTC SC is made through the metadata element "dc:subject", as shown in figure 1. Furthermore, in the OmniPaper application profile definition, the "rdfs:range" of the metadata element "dc:subject" are the IPTC Subject Codes. This means that the metadata element "dc:subject" only allows values originating from the IPTC SC (Pereira et al, 2004).

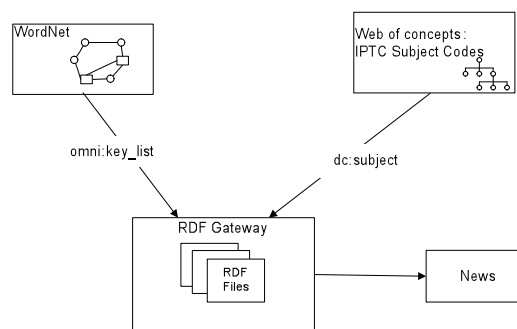


Fig. 1 OmniPaper metadata semantic layer

In addition to the navigation and browsing functionalities through the concepts represented in the IPTC Subject Codes structure, another empowering information-organization tool was included and linked to the metadatabase: WordNet® [7]. The WordNet is "an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, adjectives and adverbs are organized into synonym sets, each representing one underlying lexical concept. Different relations link the synonym sets" [7].

An RDF-encoded version of the WordNet 1.6 was downloaded and stored in a local metadatabase. Its connection with the metadata elements of the articles stored in the metadatabase is performed through the metadata element "omni-keyList" (Baptista, 2004), as shown in figure 1. This feature enables the user to perform query expansion and thus refine the search term introduced, through the semantics of query returned to the user.

The system design of the OmniPaper RDF prototype is shown in figure 2.

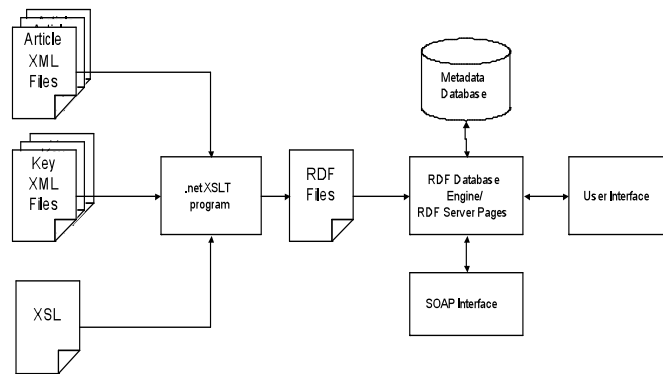


Fig. 2 OmniPaper RDF Prototype System Design

The combination of the subject thesaurus (IPTC SC) with a lexical thesaurus (WordNet), implemented in the OmniPaper RDF prototype to enhance both user queries and navigation is considered rare within RDF applications (Baptista, 2004). Moreover, the fact of these search and browsing functionalities, using a subject thesaurus and a lexical thesaurus supported by central RDF metadatabase, instead of the full text search, directly in the local archive, usually increase the response time and the results therefore have low levels of precision and recall to the user.

The different search and browsing mechanisms developed in the OmniPaper helps the user to disambiguate and refine their query. The semantic query expansion expands the words appearing in the query to semantically related words. Semantic relations are taken from the lexical thesaurus WordNet. Stemming reduces a word to its basic form (stem). Both keywords, extracted from documents, and query words have been stemmed. The query refinement tool gives suggestions to use related words or other meanings of a word appearing in the query. That way, users can get feedback on and change the semantics of their query, resulting in more accurate results.

In order to take advantage of these search and navigation functionalities and the whole RDF infrastructure developed in the OmniPaper RDF prototype, it was implemented a system which is an instance of the OmniPaper RDF prototype, in the context of scientific publication. The following sections present the work developed in the instantiation of the OmniPaper RDF prototype in the context of the scientific contents.

3 Instantiation of the OmniPaper RDF Prototype in the Context of Scientific Literature

The search mechanisms developed in the OmniPaper RDF prototype, previously presented, demonstrate that the OmniPaper RDF prototype goes far beyond the current full-text search methods. An instance of this prototype in the context of scientific literature aims to improve the daily work of scholars providing sophisticated mechanisms to access and search for new scientific contents.

The implemented system follows the concept of the OmniPaper RDF prototype. However the OmniPaper system was developed in the scope of the news published in newspapers, and its instantiation in the context of scientific publications required some changes at the level of the data and of some processes. In fact, the metadata structure was the principal change in the instantiation process of the OmniPaper RDF prototype, as a consequence of the specific features that characterizes the scientific contents. Another difference stands for the RSS approach followed in the definition of the metadata layer, while in the OmniPaper RDF prototype the metadata description followed the RDF/XML

technology. However this can not be considered as a relevant difference because the RSS version 1.0 used, conforms the W3C's RDF specification and is extensible via XML-namespace and/or RDF based modularization [1]. The extensibility mechanisms built in to RSS 1.0 ensure the semantically rich metadata description of any resource which can be identified by a URI. For this reason, the RSS 1.0 was used for metadata encoding and also because it increases interoperability with other RDF/XML applications and enables the rich semantic description of Web resources. However, in the OmniPaper RDF prototype an RDF file holds one description about one article/resource that is stored in one of the local archives, while in the RSS approach a RSS feed is defined containing a list of items with titles, links, descriptions and other terms of a set of scientific articles. The differences between the two approaches are in metadata elements used in the description of the Web resources.

Finally, the last difference is related to the navigation functionality available in both systems. In the system developed in the scope of scientific literature it was considered more adequate to use the controlled vocabulary ACM CCS instead of the IPTC Subject Codes used in the OmniPaper RDF prototype, due to the different data set used. In the OmniPaper system were used digital articles published in European newspapers, while in the instantiation system were used scientific articles. The use of the WordNet remained because it is a lexical thesaurus oriented to organize information, and therefore its use is not restricted to any specific kind of information resource.

4 Implementation of the Metadata Layer

As in the OmniPaper project, the principal aspect of the implemented system was the definition of the metadata layer. The purpose of the metadata layer is to support the search and navigation functionalities and facilitate the access to the contents presented within the feed repository. Another goal concerns the use of the central metadatabase to accomplish the syndication of contents, through the RSS approach.

Besides the RSS technology, other sophisticated syndication frameworks have been increasingly used to syndicate contents. According to Hammond "the Digital Library community has been actively deploying the Open Archives Initiative Protocol Metadata Harvesting (OAI-PMH) for syndicating metadata" (Hammond et al, 2004). Both of these frameworks are built on the same technologies, the XML, and both support the exchange of data. Although their purposes are rather different, both support various vocabularies. "RSS is predominantly used for syndicating content (usually through references to that content), while OAI-PMH is primarily focussed on harvesting metadata. RSS defines a simple encapsulation methodology that can be used by different applications (typically the RSS readers), while OAI-PMH defines both a schema and an application level protocol" (Hammond et al, 2004). RSS implement the data transfers to the user desktop, while OAI-PMH performs system-to-system processes (usually institutional repository-to-repository synchronizations) (Hammond et al, 2004).

The key difference between RSS and OAI-PMH results from the fact that RSS is mainly used in Business-to-Consumer (B2C) solution, as it enables an extensive visibility of the new contents published in the different Web pages, through the aggregation and distribution of these contents to the RSS feeds subscribers' desktop (Hammond et al, 2004).

Both of these technologies were involved in the implemented system. In fact the data set used was described in XML schema OAI-PMH, which was transformed as RSS

applications and included in a metadatabase to support the metadata search, the navigation and the metadata syndication. The description of scientific contents as RSS feeds enables scholars to subscribe them and thus be aware of any related novelties.

The typical elements of a RSS document include the metadata elements title, link and (optionally) description. However the consumer of scientific feed requires more information besides the regular metadata elements usually used in the RSS documents. In fact, the user requires more information about the article, in order to be able to cite, or produce a citation for a given article within a serial (Hammond, 2003). Consequently rich metadata was necessary in describing the article along with the article title, link and description. Thus, taking into account that the RSS 1.0 is based on RDF, makes the RSS 1.0 increasingly interesting to be used in the description of scientific publications and therefore ideally suited to the inclusion of supplementary metadata (Hammond et al, 2004). Several metadata standard vocabularies widely used in the domain of the scientific literature were analyzed in order to define a set of metadata elements which best describe the features of the scientific contents. The vocabularies were analysed in terms of the semantics of their elements, their usage in the community, and their interoperability across communities. This analysis resulted in the selection of DCMT vocabulary in the definition of the metadata structure, since it provides a variety of metadata elements that conforms totally all the necessary descriptions requirements of the information resources used, it enables the metadata interoperability, and it is an ISO and an ANSI/NISO standard. The selected elements were included in the application profile and the rules for metadata encoding were defined according to the structure defined in the RSS template. The encoding of the metadata elements followed the recommendations made by Kokkelink and Schwänzl in the document "Expressing Qualified Dublin Core in RDF/XML" [4], although it is still a DCMI candidate recommendation.

A scientific article described as an item included in a RSS feed is followed presented.

```
<item>
<title>Reconhecimento de Voz - Voice Car System (VCS)</title>
<link>http://hdl.handle.net/123456789/20</link>
<description>Os avanços tecnológicos possibilitam que os comp
a interpretem, de forma a executarem determinadas tarefas previamen
aplicabilidade destas novas tecnologias desenvolvemos uma aplicação q
interface de reconhecimento de voz, recorrendo para isso ao módulo Se
aplicação em causa simula o controlo de algumas funções de u
reconhecimento da voz humana. Numa utilização a nível real, a intera
reconhecimento da voz deverá ser efectuada através de um módulo que p
no computador de bordo do automóvel.</description>
  <dc:publisher>APSI</dc:publisher>
  <dc:creator>
    <rdf:Seq>
      <rdf:li>Neiva, Hugo</rdf:li>
      <rdf:li>Paulino, Juvenal</rdf:li>
      <rdf:li>Santos, João</rdf:li>
      <rdf:li>Silva, Bruno</rdf:li>
      <rdf:li>Silva, Pedro</rdf:li>
    </rdf:Seq>
  </dc:creator>
  <dc:rights>APSI. Todos os direitos reservados</dc:rights>
  <dc:subject rdf:datatype="http://www.acm.org/class/1998/acmccs98
  <dcterms:extent>868596</dcterms:extent>
```



```

<dcterms:medium>
  <dcterms:IMT>
    <rdf:value>pdf</rdf:value>
  </dcterms:IMT>
</dcterms:medium>
<dc:language>pt </dc:language>
<dcterms:created>
  <dcterms:W3CDTF>
    <rdf:value>2004-11-03</rdf:value>
  </dcterms:W3CDTF>
</dcterms:created>
<dcterms:issued>
  <dcterms:W3CDTF>
    <rdf:value>2005-05-26T20:09:02Z</rdf:value>
  </dcterms:W3CDTF>
</dcterms:issued>
</item>

```

Fig.3 Description of a scientific article using RSS

5 Implementation of the Conceptual Layer

The design of the system implemented in the context of scientific publications (shown in figure 4) is the same developed in the OmniPaper RDF prototype. The whole RDF model developed in the OmniPaper was instantiated in context of scientific literature. The changes performed were at the data level and in some processes.

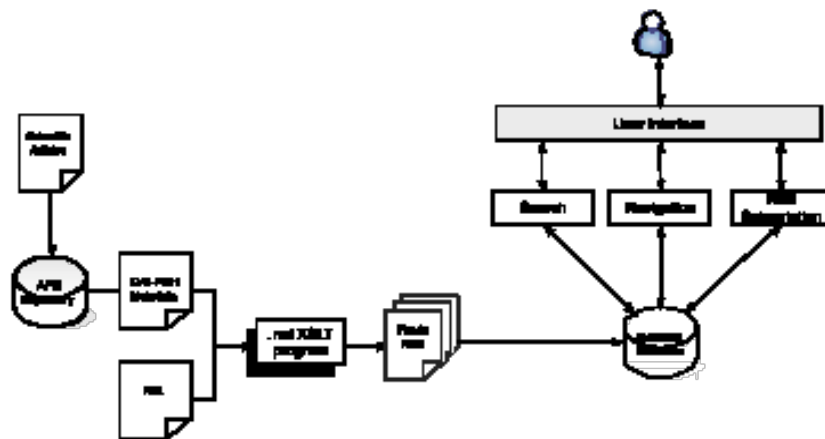


Fig. 4 System Design

The scientific contents used in the developed system are provided by the Associação Portuguesa de Sistemas de Informação (APSI) [10]. APSI has an institutional repository that stores, preserves, disseminates and enables access to the articles published in the Information Systems Journal and articles published in the APSI' conference (CAPSI) proceedings.

APSI provided the metadata information of the articles in the XML OAI-PMH format. Then, in order to get the metadata information encoded accordingly to the RSS structure defined in the template, a stylesheet to perform the transformation was developed. This process resulted in the creation of the RSS feed organized by a list of items. Each item contains the metadata description of one article. RDF triples were extracted from the RSS feed and stored in the metadatabase, through the use of the RDF Gateway, a Microsoft Windows based

native RDF database management system combined with a HTTP server. Some RDF Server Pages (RSPs) were defined in order to provide some functionality for the end user.

The metadatabase supports the navigation and browsing functionalities enabling the user to search through the metadata layer and not directly in the information source. To develop the navigation mechanism the available RDF-S version of the *ACM CCS* [11] was used. The scientific articles provided by APSI weren't classified neither using the *ACM CCS* nor any other controlled vocabulary, thus this work still had to be done manually. The connection between the *ACM CCS* subject tree to the metadatabase is performed through the "dc:subject" metadata element, as illustrated in figure 5. This fact allowed the user to subscribe the articles within a specific subject according to the specific interest areas of each subscriber.

Besides this functionality, the use of the WordNet was also instantiated from the OmniPaper RDF prototype. The use of this lexical thesaurus improves the search procedure, since it allows the relationship of the input concept with others, enabling the user to perform query expansion and thus redefine his search. The connection to the articles' descriptions stored in the metadatabase was made through the "description" metadata element, as illustrated in figure 5.

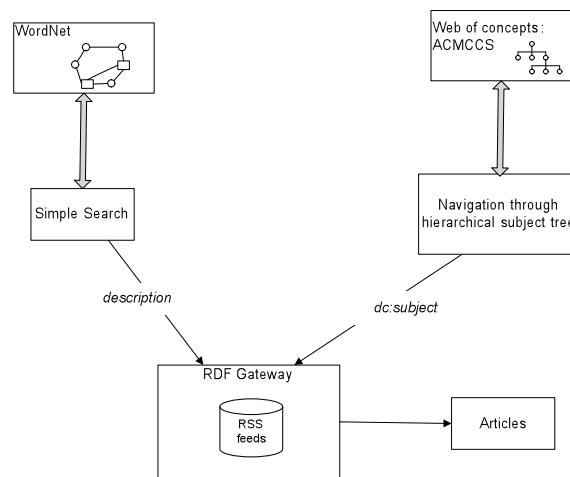


Fig. 5: Metadata semantic layer of the system

Furthermore, the metadatabase also enabled the syndication of scientific contents that are included in the RSS feed. In fact, the user doesn't need to check for new articles published in the APSI repository, because since the user subscribed the RSS feed he can be aware of new issues that have been published through an RSS Reader application the user is familiar with.

Figure 6 shows the interface of the implemented system.

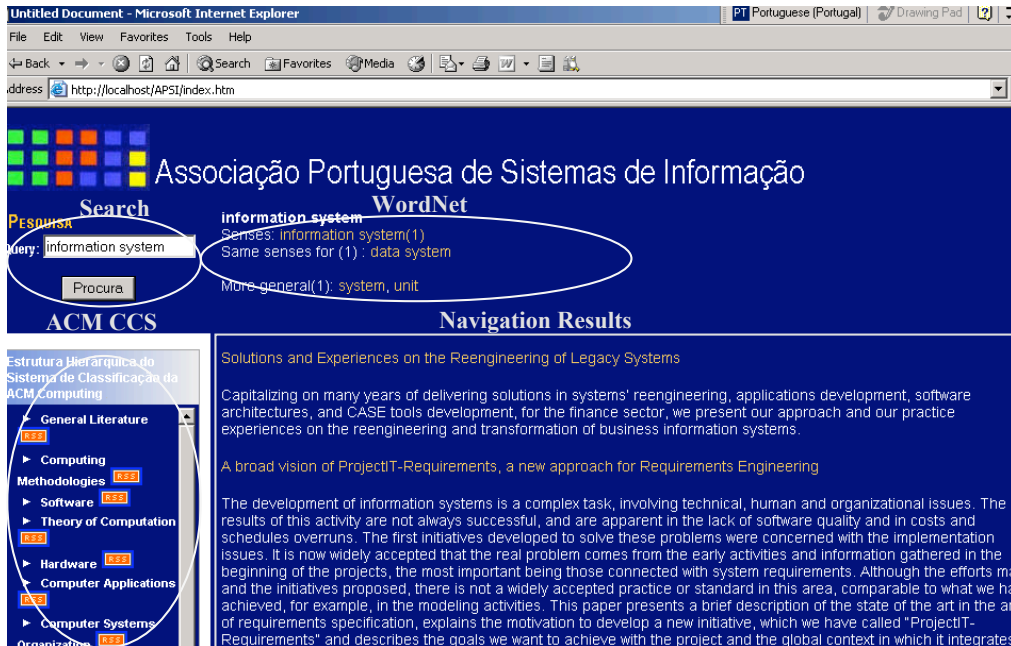


Fig. 6: Print screen of the developed prototype

6 Conclusions and Future Work

The development of information and communication technologies, and the increased use of the Internet associated to the user needs have been contributing to a deep restructure of the traditional means used in the publication of information. The RDF prototype implemented in the OmniPaper project has already proven to be an efficient system to search for news contents published in newspapers. Moreover the RDF application combines completely with subject ontology and a lexical thesaurus to enhance both user search and navigation. In this context, the instantiation of the OmniPaper RDF prototype in the scope of the scientific publication would improve the current mechanisms used to access and distribute the scientific research developments. This article described the steps followed in the instantiation of the OmniPaper RDF prototype in the context of scientific publications. The main difference between the two systems was the use of the RSS technology in the metadata description, enabling the syndication of the scientific feeds.

As future work it is necessary to evaluate the system implemented, in order to determine the relevance of the results returned and its usability. It would also be interesting to harvest several repositories to the metadatabase implemented in this system, in order to provide a more complete service with more information.

References

- Baptista, A. A. (2004) "Searching and browsing using RDF-Encoded Metadata: the case of OmniPaper", *Canadian Journal of Communication*, Vol 29 N°3, pp. 317-328. Available from: <https://repositorium.sdum.uminho.pt/handle/1822/5080>.
- Hammond, T. (2003) "Why Choose RSS 1.0?" [on-line]. *XML.com*. Available from: <http://www.xml.com/pub/a/2003/07/23/rssone.html>.

- Hammond, T., Hannay, T. and Lund, B. (2004). "The Role of RSS in Science Publishing Syndication and Annotation on the Web", *D-Lib Magazine* [on-line], Vol 10 N°12. Available from: <http://www.dlib.org/dlib/december04/hammond/12hammond.html>.
- Mika, P., Applied Ontology-based Knowledge Management: A Report on the State-of-the-Art, in Master. 2002, Vrije Universiteit: Amsterdam, from <http://www.cs.vu.nl/~pmika/thesis/pmika-thesis-full.doc>
- Paepen, B. (2005). "Blueprint: a universal standard model for efficient information retrieval". Technical Report of the OmniPaper Project, 28 February of 2005. Available from: <http://www.omnipaper.org>.
- Pereira, T. and Baptista, A. A. (2004), "Incorporating a Semantically Enriched Navigation Layer Onto an RDF Metadatabase", *Building digital bridges: linking cultures, commerce and science: Proceedings of the ICCC International Conference on Electronic Publishing, ELPUB, 8*, Brasília, Brasil, July 2004, Engelen, J., Costa Sely., M. S., Moreira, Ana Cristina S., ed. Lit. Available from: <https://repositorium.sdum.uminho.pt/handle/1822/604>
- Pereira, T. and Baptista, A. A. (2003). "The OmniPaper metadata RDF/XML prototype implementation", *ICCC International Conference on Electronic Publishing, ELPUB, 7 - From information to knowledge: proceedings*, Universidade do Minho, Guimarães, Portugal, 26 June 2003, pp. 261-268. Available from: <https://repositorium.sdum.uminho.pt/handle/1822/604>.
- Pereira, T., Yaginuma, T. and Baptista, A. A. (2003a). "OmniPaper - Arquitectura de metadados e sua implementação no RDF Gateway", *Congresso Luso-Moçambicano de Engenharia CMLE'2003, 3th*, Maputo, Moçambique, 20 August 2003, pp. 1299-1306. Available from: <http://repositorium.sdum.uminho.pt/handle/1822/280>
- Pereira, T., Yaginuma, T. and Baptista, A. A. (2003b). "Perfil de Aplicação e Esquema RDF dos Elementos de Metadados do Projecto OmniPaper *Congresso Luso-Moçambicano de Engenharia CMLE'2003, 3th*, Maputo, Moçambique, 20 August 2003, pp. 1307-1316. Available from: <http://repositorium.sdum.uminho.pt/handle/1822/281>
- Yaginuma, T., Pereira, T., Ariza, C. e Baptista, A. A., (2004). Implementation of Metadata for OmniPaper RDF Prototype. In: Sugimoto, S., ed. *International Symposium on Digital Libraries and Knowledge Communities in Networked Information Society DLKC'04, 2-5 Março 2004*, Tsukuba, Ibaraki, Japan. Available from: <https://repositorium.sdum.uminho.pt/handle/1822/606>.

Web Sites

- [1] <http://web.resource.org/rss/1.0/spec#>.
- [2] <http://www.dublincore.org/documents/dces/>.
- [3] <http://www.dublincore.org/>.
- [4] <http://www.dublincore.org/documents/2002/04/14/dcq-rdf-xml/>.

- [5] <http://www.niso.org/standards/resources/Z39-85.pdf>
- [6] <http://www.niso.org/news/releases/PRDubCr.html>
- [7] <http://www.cogsci.princeton.edu/~wn>
- [8] <http://xml.coverpages.org/NITF30-subject-codes.html>.
- [9] <http://www.iptc.org/NewsCodes/>
- [10] <http://www.apsi.pt/>
- [11] http://dspace-dev.dsi.uminho.pt:8080/pt/addon_acmccs98.jsp

About the authors

Teresa Pereira is currently an assistant lecturer at the Superior School of Business Studies of Polytechnic Institute of Viana do Castelo, She is also a Ph.D. student at the Department of Information Systems of University of Minho. She accomplished the graduation in Mathematics and Computer Science at University of Minho in 2002 and obtained the MSc degree in Information Technologies in 2006. From 2002 to 2004 she worked as a researcher in the OmniPaper project. Teresa research interests include Semantic Web, and electronic publication. She is also interested in Data Surveillance, ontologies and security information systems. Teresa Pereira is the corresponding author and can be contacted at: tpereira@esce.ipv.pt.

Ana Alice Baptista is currently assistant professor at the Department of Information Systems of University of Minho, Ana graduated as a Systems and Informatics Engineer, obtained a MSc degree in Informatics and a Doctor degree in Information Technologies and Systems. She has been publishing in the areas of Scholarly Communication, Digital Libraries and Semantic Web. She is also interested in the social aspects of the Internet, primarily on its relationships with scholarly communication.